Supporting Information

Novel Chemical Space Exploration via Natural Products

Josefin Rosén[a,*], Johan Gottfries[b], Sorel Muresan[c], Anders Backlund[a], and Tudor I. Oprea[d]

[a]*Division of Pharmacognosy, Department of Medicinal Chemistry, BMC, Uppsala University, Box 574, S-751 23 Uppsala, Sweden*
[b]*Pharmnovo Inc., Sahlgrenska Science Park, S-413 46 Gothenburg, Sweden*
[c]*DECS Global Compound Sciences, AstraZeneca R&D, S-431 83 Mölndal, Sweden*
[d]*Division of Biocomputing, MSC11 6145, University of New Mexico School of Medicine, Albuquerque, NM 87131, USA.*

[*]corresponding author: Josefin.Rosen@fkog.uu.se

Abbreviations: ACE, angiotensin-converting enzyme; A2, angiotensin II; AT1, angiotensin receptor I; CaCh, calcium channel; PPI, proton pump inhibitor; MAO, monoamine oxidase; SSRI, selective serotonin reuptake inhibitor; HIV-1, human immunodeficiency virus type 1; RT, reverse transcriptase; PR, protease; IN, integrase.

Contents:

Biological activity regions in CSSM

Considering the large number of different activities encountered in the biologically relevant chemical space, we explored if there are specific volumes of chemical space associated with certain biological activities. If true, this paradigm would enable us to partition the biologically relevant CSSM into smaller, more manageable sections linked to, or characteristic of certain activities. It is a central principle of medicinal chemistry that structurally similar molecules often have similar biological activities[1], although this has been recently questioned[2]. We recently demonstrated that the ChemGPS-NP concept is suitable for differentiation of biological activities in a study where we reported how anticancer drugs that are cytotoxic by several different mechanisms, cluster in accordance with their respective mode of action[3]. In the present paper we examine similarity based on the property neighbourhood defined by the 35 ChemGPS-NP molecular descriptors. WOMBAT compounds were mapped and labelled with regard to their tagged biological activities or targets.

Antihypertensives

Many antihypertensive drugs are used to normalize the effects of high blood pressure. Here we explore four of them: Angiotensin-converting enzyme (ACE) inhibitors lower blood pressure by reducing the production of the vasoconstrictor angiotensin II (A2) in the blood stream; angiotensin receptor 1 (AT1) antagonists work by blocking the effect of A2 on the blood vessel walls; calcium channel (CaCh) blockers affect the way calcium ions are used in blood vessels and heart muscle, and have a vasodilator effect; and $\beta 1-$adrenergic receptor ($\beta 1$) blockers lower the blood pressure by slowing the heart rate, and reducing the contractive force of the heart. The antihypertensives show clustering in accordance with these four investigated subgroups as illustrated in Figure S1A. A1 antagonists stand out from the rest by being larger, less aromatic, more polar and clearly more flexible than most other compounds. In general, CaCh blockers are the most lipophilic subgroup.

Gastric secretion inhibitors

Proton pump inhibitors (PPI) are a group of drugs causing long-lasting reduction of gastric acid production by irreversibly blocking H+/K+ ATPase, i.e. the gastric proton pump. The proton pump is responsible for secreting H+ ions into the gastric lumen. The lack of the acid in the stomach facilitates the healing of duodenal ulcers, and reduces the pain from dyspepsia. The H2 receptor antagonists are drugs used to block the action of histamine on parietal cells in the stomach, and thereby decreasing acid production by these cells. Both of the two subgroups here mentioned showed distinct clusters. PPIs were generally more aromatic than the H2 receptor antagonists (see Figure S1B).

Anti-allergy agents

Many anti-allergy agents are reversible competitive blockers of histamine at H1 receptors. As can be seen in Figure S1B the H1 receptor antagonists are tightly clustered when mapped with ChemGPS-NP. Both H1 and H2 receptor antagonists block the actions of histamine, but they can be distinguished since H1 receptor antagonists are mainly more aromatic, more lipophilic and more rigid as can be interpreted from Figure S1B.

Antidepressants

Monoamine oxidase (MAO) inhibitors are a class of antidepressant drugs prescribed for the treatment of depression acting by inhibiting the activity of MAO by preventing the breakdown of monoamine neurotransmitters. Selective serotonin re-uptake inhibitors (SSRIs) act by inhibiting the re-uptake of serotonin after being released in synapses. As a result, serotonin stays in the synaptic gap longer, and can thereby be repeatedly recognized by the receptors of the recipient cell, stimulating it. Serotonin is necessary for communication between nerve cells and it affects emotions, behaviour, mood and circadian rhythms. Both SSRI and MAO-inhibitors clustered in accordance with their respective classification (see Figure S1C), indicating that SSRIs are generally larger and more lipophilic, aromatic, and rigid than the MAO-inhibitors.

HIV inhibitors

Human immunodeficiency virus type 1 (HIV-1) infection is now pandemic. The recognition that several different classes of agents working at different places in the viral life cycle could inhibit HIV-1 replication remains the most important advance in this field, for both treatment and prevention. Antiretroviral drugs are only suppressive and to avoid that HIV-1 develops drug-resistant mutants it is essential to use a combination of drugs. The reverse transcriptase (RT) enzyme from HIV-1 is of central importance to HIV replication, and RT inhibitors have therefore become highly interesting therapeutic agents. The HIV-1 protease (PR) hydrolyzes viral poly-proteins into functional protein products that are fundamental for viral assembly and activity. The PR inhibitors utilized as drugs elicit highly specific inhibition of PR. HIV-1 also, as part of its infection cycle, depends upon the integration of a DNA copy of its viral genome into host cell chromosomes. This integration process is catalyzed by HIV-1 integrase (IN), which makes this an attractive and rational target. The substrates for RT, PR and IN clustered with their respective subgroups (see Figure S1D). Generally substrates for PR were found to be slightly larger and more lipophilic then the substrates for the other two subgroups.

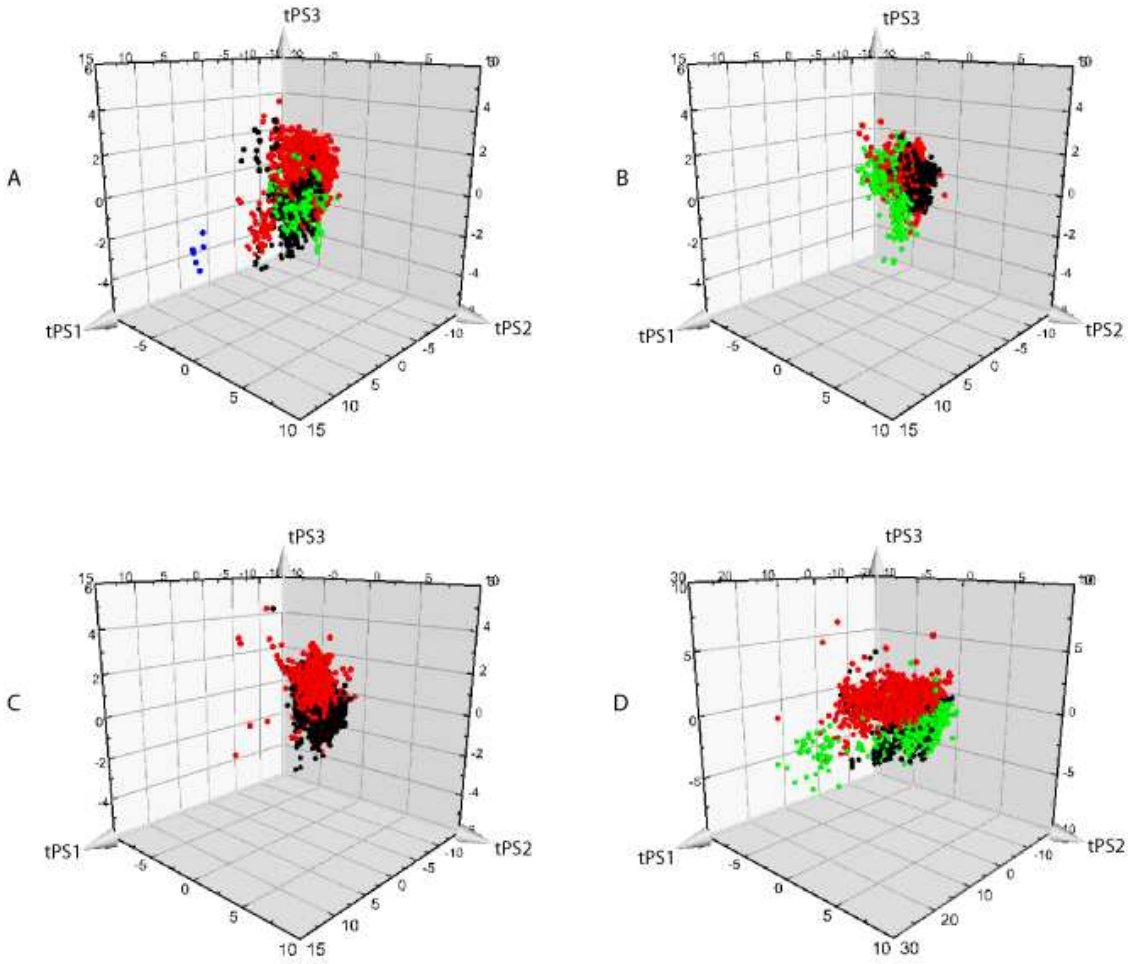

Figure S1. Predicted score plots of the first three components of ChemGPS-NP, illustrating a number of found activity regions in biologically relevant chemical space: (A) demonstrates the mapping of the antihypertensive drugs. The antihypertensives show clustering in accordance with the four investigated subgroups. ACE inhibitors are shown in black, A2 antagonists in blue, $\beta 1-$blockers in green, and CaCh antagonists in red; (B) shows the mapping of the gastric secretion inhibitors and the anti-allergy histamine H1 antagonists. Distinct clusters are obtained. PPIs are shown in black, the H2 antagonists in green, and the H1 antagonists in red; (C) shows the mapping of the antidepressant drugs. Both SSRIs and MAO inhibitors cluster with its respective classification; (D) reveals the mapping of the HIV-1 inhibiting drugs. The substrates for RT, PR, and IN clustered with their respective subgroups.

References

(1) Martin, Y. C.; Kofron, J. L.; Traphagen, L. M. Do structurally similar molecules have similar biological activity? *J. Med. Chem.* **2002,** *45*, 4350-4358.

(2) Maggiora, G. M. On outliers and activity cliffs-why QSAR often disappoints. *J. Chem. Inf. Model.* **2006,** *46*, 1535.

(3) Rosén, J.; Rickardson, L.; Backlund, A.; Gullbo, J.; Bohlin, L.; Larsson, R.; Gottfries, J. ChemGPS-NP mapping of chemical compounds for prediction of anticancer mode of action. *QSAR Comb. Sci.* **2009,** *Accepted for publication*.